PCT/AU00/01310

10/088334

REC'D 24 NOV 2000
WIPO PCT

Patent Office
Canberra

AU 00/1310

I, CASSANDRA RICHARDS, ACTING TEAM LEADER EXAMINATION SUPPORT & SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 3667 for a patent by THE UNIVERSITY OF MELBOURNE filed on 26 October 1999.

4

BEST AVAILABLE COPY

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

WITNESS my hand this
Seventeenth day of November 2000

CASSANDRA RICHARDS
ACTING TEAM LEADER
EXAMINATION SUPPORT & SALES

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

P/00/009
Regulation 3.2

AUSTRALIA
Patents Act 1990

PROVISIONAL SPECIFICATION

FOR THE INVENTION ENTITLED:

"EMPHASIS OF SHORT-DURATION TRANSIENT SPEECH FEATURES"

Applicant:

THE UNIVERSITY OF MELBOURNE

The invention is described in the following statement:

EMPHASIS OF SHORT-DURATION TRANSIENT SPEECH FEATURES

Field of the Invention

This invention relates to the processing of signals derived from sound stimuli, particularly for the generation of stimuli in auditory prostheses, such as cochlear implants and hearing aids, and in other systems requiring sound processing or encoding.

Background of the Invention

Various speech processing strategies have been developed for processing sound signals for use in stimulating auditory prostheses, such as cochlear prostheses and hearing aids. Such strategies focus on particular aspects of speech, such as formants. Other strategies rely on more general channelization and amplitude related selection, such as the Spectral Maxima Sound Processor (SMSP), strategy which is described in greater detail in Australian Patent No. 657959 by the present applicant.

A recurring difficulty with all such sound processing systems is the provision of adequate information to the user to enable optimal perception of speech in the sound stimulus.

Summary of the Invention and Object

It is an object of the present invention to provide a sound processing strategy to assist in perception of low-intensity short-duration speech features in the sound stimuli.

The invention provides a sound processing device having means for estimating the amplitude envelope of a sound signal in a plurality of spaced frequency channels, means for analyzing the estimated amplitude envelopes over time so as to detect short-duration amplitude transitions in said envelopes, means for increasing the relative amplitude of said short-duration amplitude transitions for the duration of said transitions, including means for determining the rate of change of said short-duration amplitude transitions for the whole duration of the transitions, and using this rate of change to determine the size of the increase in relative amplitude applied to said transitions.

In a preferred form, the means for determining the rate of change of said short-duration amplitude transitions determines a rate of change profile over time for each transition.

In a particularly preferred form, the faster/greater the rate of change of said short-duration amplitude transitions, the greater the increase to relative amplitude applied to said transitions. Furthermore, rate of change profiles corresponding to short-duration burst transitions receive a greater increase in relative amplitude than do profiles corresponding to onset transitions. In the present specification, a "burst transition" is understood to be a rapid increase followed by a rapid decrease in the amplitude envelope, while an "onset transition" is understood to be a rapid increase followed by a relatively constant level in the amplitude envelope.

The above defined Transient Emphasis strategy has been designed in particular to assist perception of low-intensity short-duration speech features for the severe-to-profound hearing impaired or Cochlear implantees. These speech features typically consist of: i) low-intensity short-duration noise bursts/frication energy that accompany plosive consonants; ii) rapid transitions in frequency of speech formants (in particular the 2nd formant, F2) such as those that accompany articulation of plosive, nasal and other consonants. Improved perception of these features has been found to aid perception of some consonants (namely plosives and nasals) as well as overall speech perception when presented in competing background noise.

The Transient Emphasis strategy is preferably applied as a front-end process to other speech processing systems, particularly but not exclusively, for stimulating implanted electrode arrays. The currently preferred embodiment of the invention is incorporated into the Spectral Maxima Sound Processor (SMSP) strategy, as referred to above. The combined strategy known as the Transient Emphasis Spectral Maxima (TESM) Sound Processor utilises the transient emphasis strategy to emphasise the SMSP's filter bank outputs prior to selection of the channels with the largest amplitudes.

As with most multi-channel speech processing systems, the input sound signal is divided up into a multitude of frequency channels by using a bank of band-pass filters. The signal envelope is then derived by rectifying and low-pass filtering

November 06, 2000

This is to inform you that the document requested in your order dated 10-31-2000 is a true reproduction of the official office record copy of that document:

60161492 CERTIFIED PAT APP AS FILED-EXPEDITE-PSF 1

MISSING DRAWINGS INCLUDED

☑ The enclosed Patent Application as Filed is a reproduction of the application as originally filed and has been recorded using high quality scanning or microfilm equipment. Copies of page/papers that were not scannable have not been included, nor have pages/papers received after the original filing date. Copies of these pages/papers may be ordered separately.

☐ The enclosed document is a reproduction of the best available source of the official office record copy of that document.

If you have any questions or need additional information, please contact our Customer Service Department.

Mailing Address:
U.S. Patent and Trademark Office
Office of Public Records, Customer Service
Crystal Gateway 4, Suite 300
Washington DC 20231

Delivery Address:
U.S. Patent and Trademark Office
Office of Public Records, Customer Service
1213 Jefferson Davis Highway, Suite 300
Arlington VA 22202

For faster processing of new orders, please specify as appropriate:
Box 9 (Copy Sales) for Uncertified copies, or Box 10 for Certified copies of PTO Documents

Voice: (703) 308-9726 Fax: (703) 308-7048 E-Mail: PTCS@USPTO.GOV or Certdiv@USPTO.GOV

Ref:HP 325501

the signal in these bands. Emphasis of short-duration transitions in the envelope signal for each channel is then carried out. This is done by: i) detection of short-duration (approximately 10 to 60 milliseconds) amplitude variations in the channel envelope typically corresponding to speech features such as noise bursts, formant transitions, and voice onset; and ii) increasing the signal gain during these periods. The gain is adjusted in proportion to the derivative with respect to time (gradient) of the signal envelope (or some similar rule, as described below in the Detailed Description section).

During periods of steady state or relatively slow varying levels in the envelope signal (over a period of approximately 60ms) no gain is applied. During periods where short-duration transition in the envelope signal are detected, the amount of gain applied can typically vary from 0dB to 12dB. The gain varies depending of the nature of the short-duration transition which can be classified as either of the following. i) A rapid increase followed by a decrease in the signal envelope (over a period of no longer than approximately 60ms). This typically corresponding to speech features such as the noise-burst in plosive consonant or the rapid frequency shift of a formant in a consonant-to-vowel or vowel-to-consonant transition. ii) A rapid increase followed by relatively constant level in the signal envelope which typically corresponds to speech features such as the onset of voicing in a vowel. Short duration speech features classified according to i) are considered to be more important to perception than those classified according to ii) and thus receive relatively twice as much gain. Note, a relatively constant level followed by a rapid decrease in the signal envelope which corresponds to abruption of voicing/sound does not receive any gain.

Brief Description of the Drawings

In order that the invention may be more readily understood, one presently preferred embodiment of the invention will now be described with reference to the accompanying drawings in which:

Figure 1 is a schematic representation of the signal processing applied to the sound signal in accordance with the present invention, and

Figures 2 and 3 are comparative electrodograms of sound signals to show the effect of the invention.

Description of Preferred Embodiment

Referring to Figure 1, the presently preferred embodiment of the invention is described with reference to its use with the SMSP strategy. As with the SMSP strategy, electrical signals corresponding to sound signals received via a microphone 1 and pre-amplifier 2 are processed by a bank of N parallel filters 3 tuned to adjacent frequencies (typically N = 16). Each filter channel includes a band-pass filter 4, then a rectifier 5 and low-pass filter 6 to provide an estimate of the signal amplitude (envelope) in each channel. In this embodiment a Fast Fourier Transform (FFT) implementation of the filter bank is employed. The outputs of the N-channel filter bank are modified by the transient emphasis algorithm 7 (as described below) prior to further processing by the SMSP strategy.

The envelope signal in each of the N channels (denoted S_n where the subscript n refers to the channel number) is further low-pass filtered 8 (denoted E_n) so as to attenuate any frequency components above approximately 50-100Hz (such as the voicing frequency). This low-pass filter (2nd order low-pass cut-off freq of 25Hz) introduces a group delay (T) to the signal (typically T = 10 ms). A running history, which spans a time period of $6 \times T$ (typically 60 ms), of the original envelope signal (denoted $S_n(t)$ as a function of time) and the low-pass filtered envelope signal ($E_n(t)$ as a function of time) is maintained 9. Time (t) is relative to the original envelope signal $S_n(t)$. An emphasis gain coefficient (denoted G_n) is then calculated 10 for each channel from the low-pass filtered envelope signal according to equation (1).

Equation. 1. $G_n = (2 \times E_n(t_1) - 2 \times E_n(t_2) - E_n(t_0)) / (E_n(t_0) + E_n(t_1) + E_n(t_2))$

Where $t_0 = -T$, $t_1 = -3 \times T$, $t_2 = -5 \times T$

(typically $t_0 = -10\text{ms}$, $t_1 = -30\text{ms}$, $t_2 = -50\text{ms}$)

The amount of gain (G_n) applied (as per equation (1)) thus varies with the behavior of $E_n(t)$ such that a rapid increases followed by a rapid decrease (over a time period of no longer than approximately $6 \times T$) in the envelope signal ($E_n(t)$) will produce the greatest values of G_n. Typically for speech-like signals, G_n can be

expected to range from 0.0 to 2.0. A rapid increase followed by a relatively constant level in the envelope signal will produce lower levels of G_n (approximately half that of the previous condition). A relatively steady-state or slow varying envelope signal will produce a negative value of G_n. So too will a steady-state level followed by a rapid decrease in the envelope signal. The emphasis gain coefficient is therefore limited 11 such that it can never fall below zero as per equation (2). Also an upper gain limit (L_n) is included to restrict the value G_n in cases where short-duration transients in the envelope signal exceed the dynamics of speech-like signals (such as the slamming of a door, or banging of a drum). L_n can be specified independently for each frequency channel (typically $L_n = 2$ for all n).

Equation. 2. $0 \le G_n \le L_n$

That is, If ($G_n > L_n$) then $G_n = L_n$

If ($G_n < 0$) then $G_n = 0$;

Where L_n = emphasis gain limit (typically $L_n = 2$)

It should be noted that whilst equations (1) and (2) define the rules used for calculation of the emphasis gain coefficient in this embodiment of the algorithm, other gradient type rules may also be applicable.

Each emphasis gain coefficient (G_n) is then used to scale 13 the original envelope signal ($S_n(t)$) according to equation (3). An emphasis gain modifier constant (K_n) has been included 12 to allow for adjustment of the overall gain of the algorithm. K_n can be specified independently for each channel so that the amount of gain applied can be adjusted for each frequency channels (typically $K_n = 2$ for all n). During periods of no short-duration transitions in the envelope signal, the emphasis gain coefficient (G_n) is equal to zero and thus $S'_n(t_l) = S_n(t_l)$. During periods when short-duration transients in the envelope signal take place, G_n will be greater than zero and additional gain to $S_n(t_l)$ is applied.

Equation. 3. $S'_n(t_l) = S_n(t_l) \times (1 + K_n \times G_n)$

where $t_l = -3 \times T$, (typically $t_l = -30\text{ms}$)

and K_n = emphasis gain modifier constant (typically $K_n = 2$)

Note that the gain coefficient (G_n) is applied to the corresponding envelope signal ($S_n(t)$) at a point in time mid-way through the analysis history (typically -30ms).

Thus an overall delay of 3 × T (typically 30ms) occurs between time from input to output of the transient emphasis algorithm. The above procedure is conducted independently for each of the channels (that is, for n = 1 to number of channels (N)). It is repeated at regular time intervals such that the gain applied per channel varies over time.

The modified envelope signal $S'_n(t)$ replace the original envelope signal $S_n(t)$ derived from the N channel filter bank 3 of the SMSP strategy. As with the SMSP strategy, M of the N channels of $S'_n(t)$ having the largest amplitude at a given instance in time are selected 14 (typically M = 6). This occurs at regular time intervals and for the transient emphasis strategy is typically 2.5ms. The M selected channels are then used to generate M electrical stimuli 15 corresponding in stimulus intensity and electrode number to the amplitude and frequency of the M selected channels (as per the SMSP strategy). These M stimuli are transmitted to the Cochlear implant 17 via a radio-frequency link 16 and are used to activate M corresponding electrode sites.

To illustrate the effect of the TESM strategy, "electrodogram" recordings from the output of a speech processor used in a cochlear implant system have been provided for the SMSP and TESM strategies (refer to Figures 2 & 3 respectively). Electrograms are somewhat similar to spectrograms for acoustic signals, and show how the excitation on each of the intra-cochlear electrodes varies with time. Time is shown along the abscissa and electrode number along the ordinate. For each stimulus pulse recorded from the output of the processor, a vertical bar is shown in the electrodogram at the time and electrode position of the stimulus. The height of the bar represents the stimulus intensity (log current) where zero-height corresponds to the hearing threshold for electrical stimulation at the specified electrode position, and maximum-height corresponds to maximum comfortable loudness for electrical stimulation at the specified electrode position. The speech token presented in these recordings was 'aka' and was spoken by a male speaker.

Inspecting the electrograms it can be seen that the amplitude of the stimuli representing the plosive burst in the medial consonant 'k' have been emphasised (ie, temporarily increased in amplitude) by the TESM strategy 18 relative to the

standard SMSP strategy. So too have the 2nd formant (F2) transition from initial vowel 'a' to 'k' 19, and from 'k' to final vowel 'a' 20. Also worth noting is that the onset of the initial vowel 'a' has been slightly emphasised 21.

Since modifications within the spirit and scope of the invention may be readily effected by persons skilled in the art, it is to be understood that the invention is not limited to the particular embodiment described, by way of example, hereinabove.

DATED: 26 October 1999

CARTER SMITH & BEADLE

Patent Attorneys for the Applicant:

THE UNIVERSITY OF MELBOURNE

Figure1.

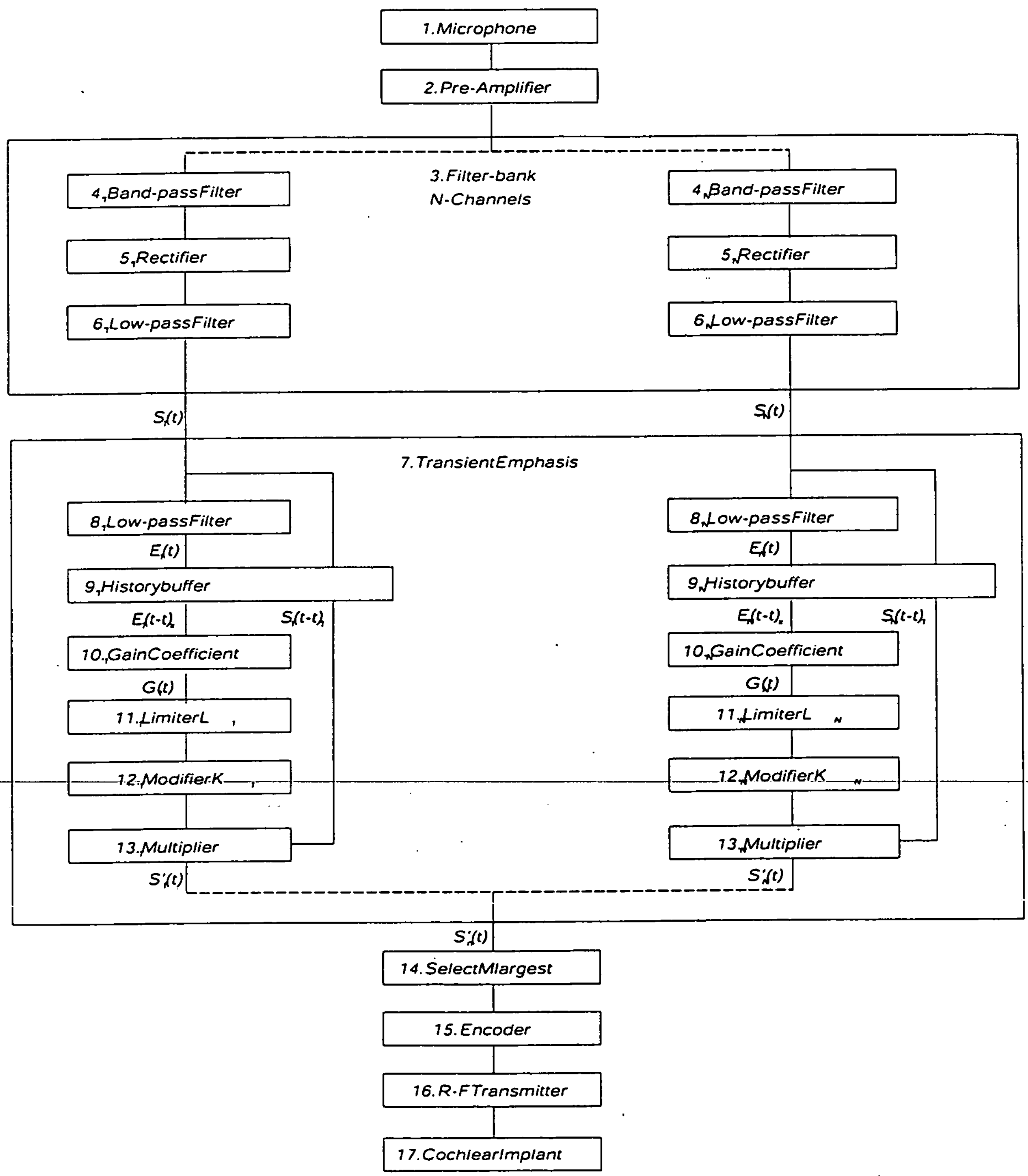
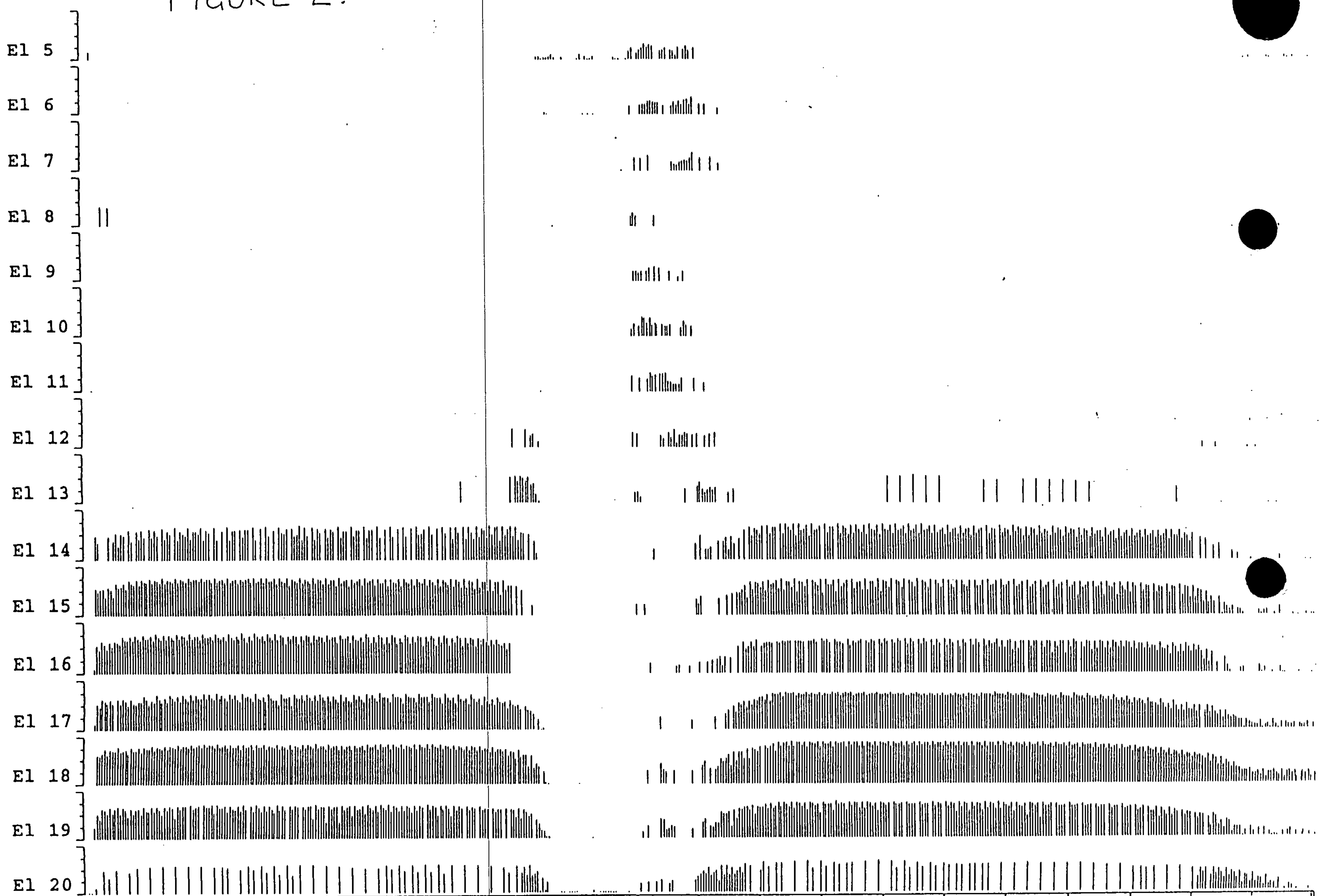
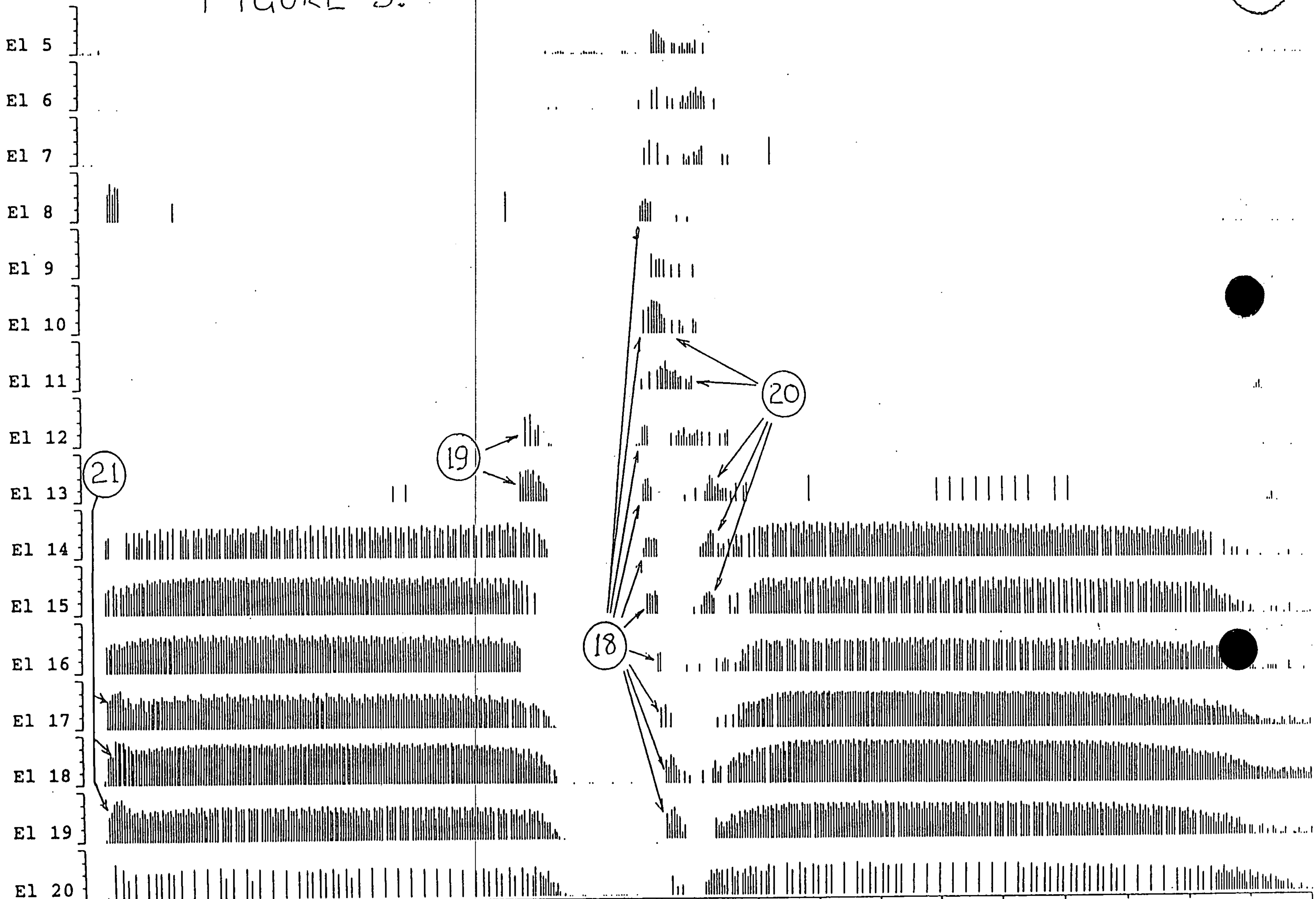

1.Microphone
2.Pre-Amplifier
3.Filter-bank
N-Channels
4,Band-passFilter
4NBand-passFilter
5,Rectifier
5NRectifier
6,Low-passFilter
6NLow-passFilter
S(t)
S(t)
7.TransientEmphasis
8,Low-passFilter
8NLow-passFilter
E(t)
E(t)
9,Historybuffer
9NHistorybuffer
E(t-t)
S(t-t)
E(t-t)
S(t-t)
10,GainCoefficient
10NGainCoefficient
G(t)
G(t)
11.LimiterL
11NLimiterL
12.ModifierK
12NModifierK
13.Multiplier
13NMultiplier
S'(t)
S'(t)
S'(t)
14.SelectMlargest
15.Encoder
16.R-FTransmitter
17.CochlearImplant

FIGURE 2.



Time 60.0 ms/div until 0.0 ms before end of recording; total 1200.0 ms.
Current levels (T-C) vs. time; File: AKA2.DF; ID: smsp aka

FIGURE 3.



Time: 60.0 ms/div until 0.0 ms before end of recording; total 1200.0 ms.
Current levels (T-C) vs. time: File: AKA1.DF; ID: newtesm2 aka

AV_LEN=8 Gen=3

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- [ ] BLACK BORDERS
- [ ] IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- [x] FADED TEXT OR DRAWING
- [x] BLURRED OR ILLEGIBLE TEXT OR DRAWING
- [ ] SKEWED/SLANTED IMAGES
- [ ] COLOR OR BLACK AND WHITE PHOTOGRAPHS
- [ ] GRAY SCALE DOCUMENTS
- [x] LINES OR MARKS ON ORIGINAL DOCUMENT
- [ ] REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- [ ] OTHER: ______________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.